

B11109014 四資管三 周詠熙

作業八：專案規劃

組名：無病不知

組員：

B11109050 洪熒孺（專案經理）

B11109014 周詠熙（技術專家）← 這我

B11109008 黃湓璇（技術專家/設計專家）

B11109046 徐名巖（研究員）

專案概述：使用 python 建立一個推薦病患看診科室的網站。

本人負責之部分：

1. 爬蟲收集資料
2. 判斷科室

以下進行詳述

爬蟲收集資料

經本組篩選後，決定爬的網站為 udn 的疾病百科，因為此網站幾乎涵蓋了所有疾病，且每個疾病的書寫格式相同，不太會有因為書寫格式不同要特別處理的情況。此外，網站上的內容發布都有專業的醫生進行審稿，避免了資料有誤的問題。主要抓取資料為各個疾病之症狀、併發症、科別、部位。

本來也有想過要使用維基百科的疾病列表，但是發現有些疾病只有英文或其他語言，中文尚未建檔，會有資料不完整之問題。而且因為維基百科是多人編輯，無法保證絕對正確，書寫格式也各不相同，爬蟲難度高。

網站連結：

UDN 疾病百科 https://health.udn.com/health/disease_list

維基百科疾病列表

<https://zh.m.wikipedia.org/wiki/%E7%96%BE%E7%97%85%E5%88%97%E8%A1%A8>

我的流程步驟為爬蟲→數據處理後存成 csv 檔案以供後續使用→使用 github actions 自動化爬蟲流程，定期更新資料（如有時間）

期間會使用到的技術為：

1. 爬蟲技術：

- Selenium：處理需要動態載入的網頁內容，模擬瀏覽器操作，如滾動、點擊等，以獲取完整的網頁資料。
- BeautifulSoup：從 Selenium 獲得的網頁源代碼中進行 HTML 標籤解析，提取所需資料。

2. 數據處理與存儲：

- Pandas：用於資料清理和格式整理，將數據轉換為 CSV 格式存檔，方便後續使用。
- 正則表達式 (re)：用來從非結構化的文字中擷取關鍵資料，提高資料一致性。

3. 自動化和排程：

- GitHub Actions：設定自動化流程，按指定時間執行 Selenium 腳本，定期更新資料。

判斷科室

判斷科室的部分，目前主要的想法是將抓到的症狀分類（目前是打算根據部位，但是可能會有一個疾病有多個部位的問題）。使用者可以在前端點擊目前的症狀，根據點選的症狀逐步顯示可能的疾病及推薦的科室。

大概會有以下幾個 class：

1. Symptom

- 說明：此 class 用於表示每個症狀
- attribute：
 - name：症狀名稱（例如：「咳嗽」、「腹痛」）。
 - related_body_parts：與該症狀相關的部位（如「胸部」、「腹部」）。
- function：
 - get_related_diseases()：根據症狀查詢相關的疾病。

2. BodyPart

- 說明：此 class 用於表示身體的各個部位，以便根據部位分類症狀及對應疾病。
- attribute：
 - name：部位名稱（例如：「頭部」、「胸部」、「腹部」）。
 - related_symptoms：此部位常見的症狀列表。
 - related_departments：推薦的科室（例如：「耳鼻喉科」、「神經內科」）。
- function：
 - get_recommended_departments()：返回該部位常見症狀所推薦的科室。

3. Department

- 說明：此 class 用於表示醫院的各個科室，例如「內科」、「皮膚科」、「骨科」等。
- attribute：
 - name：科室名稱。
 - specialized_body_parts：科室主要負責的身體部位
 - specialized_symptoms：科室擅長處理的症狀列表。
- function：
 - get_recommended_diseases(symptom)：根據症狀推薦可能的疾病。

4. Disease

- 說明：描述疾病的 class，用於儲存各個疾病的詳細資料。
- attribute：
 - name：疾病名稱。
 - symptoms：疾病的主要症狀列表。
 - complications：可能的併發症。
 - related_departments：推薦的科室列表。
 - affected_body_parts：疾病主要影響的身體部位。
- function：
 - get_recommended_department()：根據症狀和部位推薦科室。
 - get_symptom_match(symptoms)：用於判斷該疾病是否符合給定的症狀。

5. SymptomAnalyzer

- 說明：症狀分析工具，用於根據使用者輸入的症狀來逐步篩選可能的疾病和推薦科室。
- attribute：
 - selected_symptoms：使用者選擇的症狀列表。
- function：
 - add_symptom(symptom)：增加使用者選擇的症狀。
 - analyze_possible_diseases()：根據症狀匹配可能的疾病。
 - recommend_departments()：推薦相關的科室。